2.1 - Introduction

- How do we link sample spaces and events to data?
  ↳ The link is provided by the concept of random variable

Definition: A random variable is a mapping*

$$X : \Omega \to \mathbb{R}$$

that assigns a real number $X(\omega)$ to each outcome ω.

* Technically, a random variable must be measurable.

2.13 Appendix

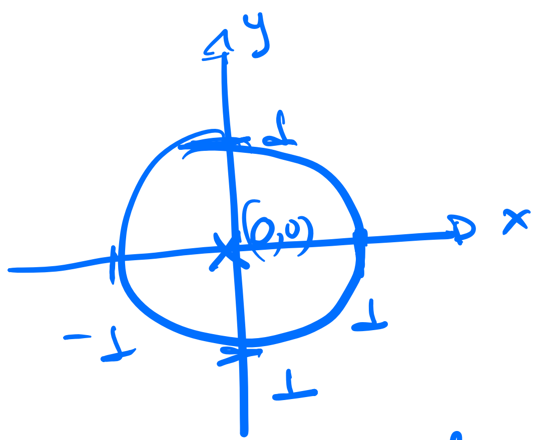
Recall that a probability measure $\mathbb{P}$ is defined on a σ-field $\mathcal{A}$ of a sample space Ω. A random variable X is a **measurable** map $X : \Omega \to \mathbb{R}$. Measurable means that, for every x, $\{\omega : X(\omega) \leq x\} \in \mathcal{A}$.

eg: Flip a coin 10 times. Let $X(\omega)$ be the number of H in the sequence ω.
If $\omega = HHTHHTHHTT$, then $X(\omega) = 6$

eg 2

Let $\Omega = \{ (x,y) ; x^2 + y^2 \le 1 \}$

↳ unit circle



* Consider drawing a point at random from Ω.

↳ Typical outcome will be $\omega = (x, y)$

examples of random variable:

$X(\omega) = x$

$Y(\omega) = y$

$Z(\omega) = x + y$

$W(\omega) = \sqrt{x^2 + y^2}$

Given a random variable X and a subset A of the real line, define

$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$, let

* $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$

* $\mathbb{P}(X = x) = \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega ; X(\omega) = x\})$

* <u>Notice</u>: That <u>X</u> denotes the random variable

& <u>x</u> denotes a particular value of <u>X</u>

2.4 Example. Flip a coin twice and let X be the number of heads. Then, $\mathbb{P}(X = 0) = \mathbb{P}(\{TT\}) = 1/4$, $\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = 1/2$ and $\mathbb{P}(X = 2) = \mathbb{P}(\{HH\}) = 1/4$. The random variable and its distribution can be summarized as follows:

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
TT	1/4	0
TH	1/4	1
HT	1/4	1
HH	1/4	2

x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

Try generalizing this to n flips. ∎

* Important example since it shows the difference mentioned above in <u>red</u>

2.2 - Distribution functions and Probability Functions

- Given a random Variable X, we define the cumulative distribution function [CDF]

<u>Definition</u>: The cumulative distribution function, or CDF, is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x)$$

example

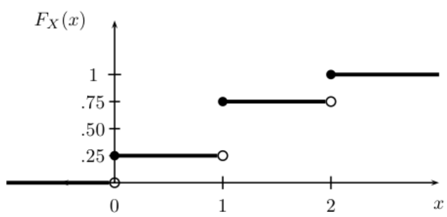


FIGURE 2.1. CDF for flipping a coin twice (Example 2.6.)

We will see later that the CDF effectively contains all the information about the random variable. Sometimes we write the CDF as F instead of F_X.

2.6 Example. Flip a fair coin twice and let X be the number of heads. Then $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ and $\mathbb{P}(X = 1) = 1/2$. The distribution function is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

* Right-continuous ??

The CDF is shown in Figure 2.1. Although this example is simple, study it carefully. CDF's can be very confusing. Notice that the function is right continuous, non-decreasing, and that it is defined for all x, even though the random variable only takes values 0, 1, and 2. Do you see why $F_X(1.4) = .75$? ∎

The following result shows that the CDF completely determines the distribution of a random variable.

2.7 Theorem:

Let X have CDF F and let Y have CDF G.
If $F(x) = G(x)$ for all x, then
$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for all A.
[for every measurable event A]

2.8 Theorem

A function F mapping the real line to $[0,1]$ is a CDF for some probability $\mathbb{P}$ if & only if F satisfies the following conditions:

1) F is non-decreasing: $x_1 < x_2$ implies $F(x_1) \leq F(x_2)$

2) F is normalised:
$$\lim_{x \to -\infty} F(x) = 0 \quad \& \quad \lim_{x \to \infty} F(x) = 1$$

3) F is right-continuous: $F(x) = F(x^+)$ for all x, where
$$F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y)$$

Proof

Suppose that F is a CDF. Let us show that 3 holds

→ Let x be a real number and let $y_1, y_2 \ldots$ be a sequence of real numbers such that $y_1 > y_2 > y_3 \ldots$, and $\lim_i y_i = x$

→ def $A_i = (-\infty, y_i]$ & $A = (-\infty, x]$

Note that $A = \bigcap_{i=1}^{\infty} A_i$ & $A_1 \supset A_2 \supset A_3 \ldots$

monotone decreasing

$\therefore \lim_i \mathbb{P}(A_i) = \mathbb{P}(\bigcap_i A_i)$

$\Leftrightarrow$

$F(x) = \mathbb{P}(A) = \mathbb{P}(\bigcap_i A_i) = \lim_i \mathbb{P}(A_i)$

$= \lim_i F(y_i) = F(x^+)$

* Note: Showing (1) & (2) is similar.

* **Definition:**

X is discrete if it takes countably* many values $\{x_1, x_2 \dots\}$.

We define the probability function or probability mass function for X by

$f_X(x) = \mathbb{P}(X = x)$

$\therefore f_X(x) \geq 0$ for all $x \in \mathbb{R}$ and

$\sum_i f_X(x_i) = 1$

$\Rightarrow$ The CDF of X is related to f_X by

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

*Note: A set is countable if it is finite or can be put in a one-to-one correspondence with the integers. The even numbers, the odd numbers and the rationals are countable; the set of real numbers between 0 & 1 is not countable.

↳ Meaning that we cannot map every number of that set on a different Natural number
[Cantor's diagonal argument]
This is also because the enumeration between real number of 0 – 1

$1/3 = 0.333 \rightarrow$ repeating.

* Definition:

A random variable X is continuous if there exist a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ & for every $a \leq b$

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\,dx$$

The function f_X is called the probability density function (PDF)

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

and $f_X(x) = F_X'(x)$ at all points x at which F_X is differentiable

* Note: $\int f(x)\,dx = \int f = \int_{-\infty}^{\infty} f(x)\,dx$

2.12 Example. Suppose that X has PDF

$$f_X(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $f_X(x) \geq 0$ and $\int f_X(x)dx = 1$. A random variable with this density is said to have a Uniform (0,1) distribution. This is meant to capture the idea of choosing a point at random between 0 and 1. The CDF is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

See Figure 2.3. ■

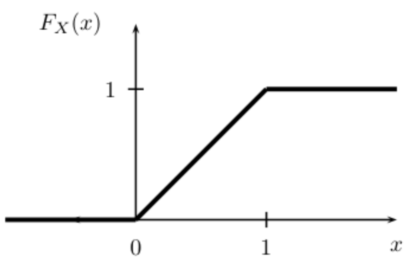


FIGURE 2.3. CDF for Uniform (0,1).

2.13 Example. Suppose that X has PDF

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{(1+x)^2} & \text{otherwise.} \end{cases}$$

Since $\int f(x)dx = 1$, this is a well-defined PDF. ■

Warning! Continuous random variables can lead to confusion. First, note that if X is continuous then $\mathbb{P}(X = x) = 0$ for every x. Don't try to think of $f(x)$ as $\mathbb{P}(X = x)$. This only holds for discrete random variables. We get probabilities from a PDF by integrating. A PDF can be bigger than 1 (unlike a mass function). For example, if $f(x) = 5$ for $x \in [0, 1/5]$ and 0 otherwise, then $f(x) \geq 0$ and $\int f(x)dx = 1$ so this is a well-defined PDF even though $f(x) = 5$ in some places. In fact, a PDF can be unbounded. For example, if $f(x) = (2/3)x^{-1/3}$ for $0 < x < 1$ and $f(x) = 0$ otherwise, then $\int f(x)dx = 1$ even though f is not bounded.

eg let $f(x) = \begin{cases} 0 & \text{for } x<0 \\ \frac{1}{(x+1)} & \text{otherwise} \end{cases}$

This is not a PDF since

$$\int f(x)\,dx = \int_0^\infty \frac{1}{x+1}\,dx = \int_1^\infty \frac{du}{u}$$

$$= \log(\infty) = \infty$$

let $u = x+1$

$\frac{du}{dx} = 1$

2.15 Lemma. *Let F be the* CDF *for a random variable X. Then:*

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$ *where* $F(x^-) = \lim_{y \uparrow x} F(y)$*;*

---

2. $\mathbb{P}(x < X \le y) = F(y) - F(x)$*;*

3. $\mathbb{P}(X > x) = 1 - F(x)$*;*

4. *If X is continuous then*

$$\begin{aligned} F(b) - F(a) &= \mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) \\ &= \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X \le b). \end{aligned}$$

It is also useful to define the inverse CDF (or quantile function).

2.16 Definition. *Let X be a random variable with* CDF *F. The* **inverse CDF** *or* **quantile function** *is defined by*[4]

$$F^{-1}(q) = \inf\Big\{x : \; F(x) > q\Big\}$$

for $q \in [0,1]$. If F is strictly increasing and continuous then $F^{-1}(q)$ is the unique real number x such that $F(x) = q$.

We call $F^{-1}(1/4)$ the **first quartile**, $F^{-1}(1/2)$ the **median** (or second quartile), and $F^{-1}(3/4)$ the **third quartile**.

Two random variables X and Y are **equal in distribution** — written $X \stackrel{d}{=} Y$ — if $F_X(x) = F_Y(x)$ for all x. This does not mean that X and Y are equal. Rather, it means that all probability statements about X and Y will be the same. For example, suppose that $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. Let $Y = -X$. Then $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ and so $X \stackrel{d}{=} Y$. But X and Y are not equal. In fact, $\mathbb{P}(X = Y) = 0$.

2.3 - Important Discrete Random Variables

*** Notation Warning ***

$X \sim F$ → indicates that X has distribution F

• POINT MASS DISTRIBUTION

X has a point mass distribution at a, $X \sim \delta_a$, if $P(X=a) = 1$

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$$

∴ probability mass function

$f(x) = 1$ for $x = a$

& 0 otherwise

• DISCRETE UNIFORM DISTRIBUTION

Let $k > 1$ be a given integer. Suppose that X has probability mass function given by

$$f(x) = \begin{cases} 1/k & \text{for } x = 1, \dots, k \\ 0 & \text{otherwise} \end{cases}$$

*We say that X has uniform distribution on $\{1, \dots, k\}$

- **BERNOULLI DISTRIBUTION**

Let X represent a binary coin flip.
Then $P(X=1)=p$ & $P(X=0)=1-p$ for some $p \in [0,1]$

→ We say that X has a Bernoulli distr.
written $X \sim Bernoulli(p)$

The probability function is
$f(x) = p^x(1-p)^{1-x}$ for $x \in \{0,1\}$

- **BINOMIAL DISTRIBUTION**

Suppose we have a coin which falls heads up with probability p for some $0 \leq p \leq 1$

Flip coin n times & let X be the # of H.

* Assume coin tosses are independent.

Let $f(x) = P(X=x)$ be the mass function

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

→ Written as $X \sim Binomial(n,p)$

Note: If $X_1 \sim \text{Binomial}(n_1, p)$ &
$X_2 \sim \text{Binomial}(n_2, p)$
$\Rightarrow X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$

Warning! Let us take this opportunity to prevent some confusion. X is a random variable; x denotes a particular value of the random variable; n and p are **parameters**, that is, fixed real numbers. The parameter p is usually unknown and must be estimated from data; that's what statistical inference is all about. In most statistical models, there are random variables and parameters: don't confuse them.

- GEOMETRIC DISTRIBUTION

X has a geometric distribution with parameters $p \in (0,1)$, written as $X \sim \text{Geom}(p)$ if

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1$$

$$\Rightarrow \sum_{k=1}^{\infty} \mathbb{P}(X = k) = p \sum_{k=1}^{\infty} (1-p)^{k} = \frac{p}{1-(1-p)} = 1$$

Think of X as the number of flips needed until the first head when flipping a coin

- POISSON DISTRIBUTION

X has Poisson distribution with parameter λ, written as $X \sim \text{Poisson}(\lambda)$ if

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x \geq 0$$

Note:

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

. The poisson is often used as a model for counts of rare events, like radioactive decay and traffic accidents.

Note: If $X_1 \sim \text{Poisson}(\lambda_1)$ &

$X_2 \sim \text{Poisson}(\lambda_2)$

$\Rightarrow X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$

Note: pg 16 of notes very important

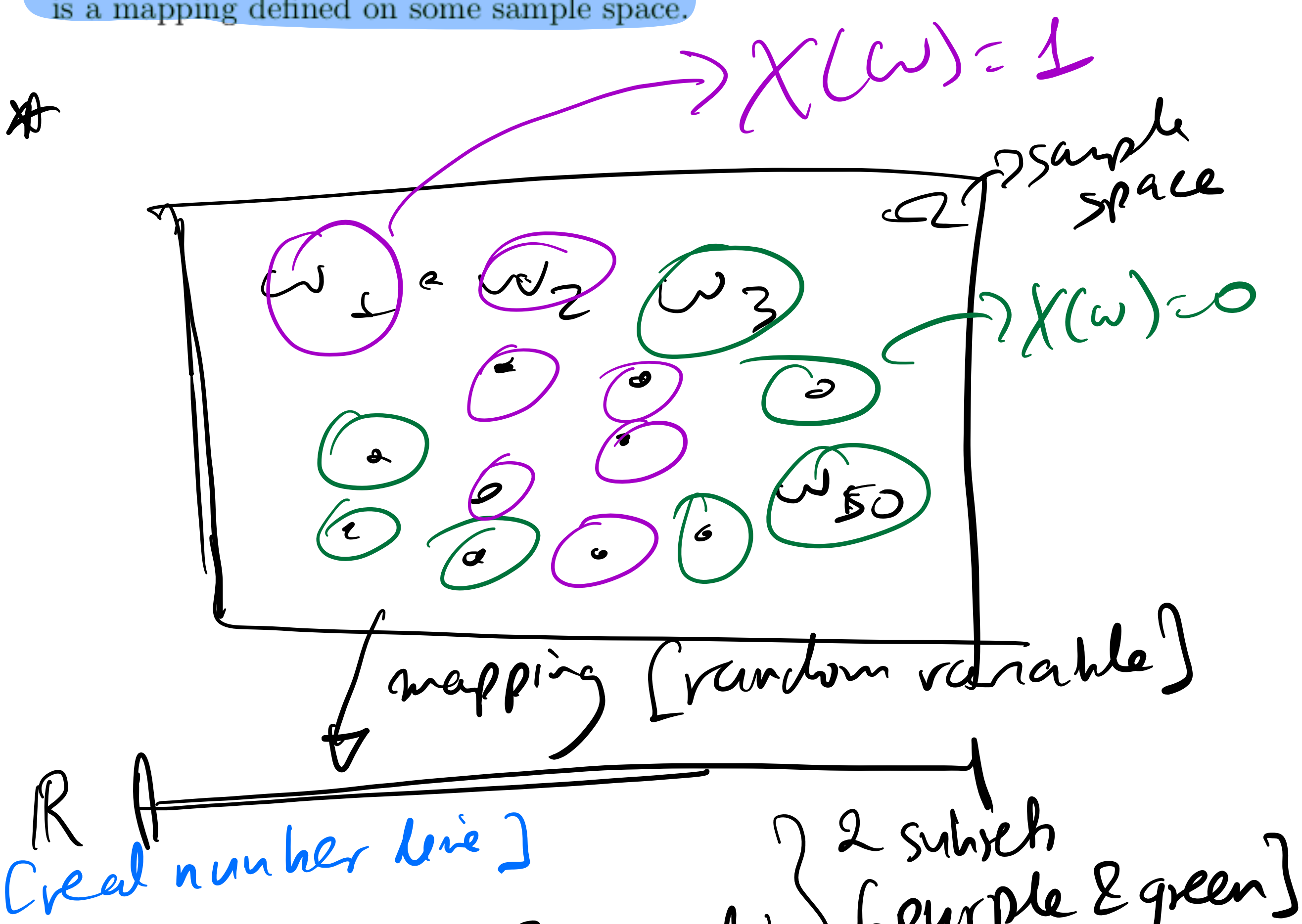
Warning! We defined random variables to be mappings from a sample space Ω to $\mathbb{R}$ but we did not mention the sample space in any of the distributions above. As I mentioned earlier, the sample space often "disappears" but it is really there in the background. Let's construct a sample space explicitly for a Bernoulli random variable. Let $\Omega = [0,1]$ and define $\mathbb{P}$ to satisfy $\mathbb{P}([a,b]) = b - a$ for $0 \le a \le b \le 1$. Fix $p \in [0,1]$ and define

x what is prob of purple event

$$X(\omega) = \begin{cases} 1 & \omega \le p \\ 0 & \omega > p. \end{cases}$$

* Probability are defined in events

Then $\mathbb{P}(X = 1) = \mathbb{P}(\omega \le p) = \mathbb{P}([0,p]) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Thus, $X \sim \text{Bernoulli}(p)$. We could do this for all the distributions defined above. In practice, we think of a random variable like a random number but formally it is a mapping defined on some sample space.



* In this case Bernoulli
X(ω) → 0, 1

2 subsets [purple & green] [events]
∴ can assign prob

2.4 – Important Continuous Random Variables

- **UNIFORM DISTRIBUTION**

Written as $X \sim \text{Uniform}(a, b)$

if $f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$

where $a < b$

→ The distribution function is:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

· NORMAL (GAUSSIAN)

X has a normal (gaussian) distribution with parameters μ & σ

written as $X \sim N(\mu, \sigma^2)$ if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$

* The parameter μ (mean)
* The parameter σ (spread) (standard deviation)

Note: Later on we will discuss Central Limit Theorem
↳ sum of random variables can be approximated by a Normal distribution

→ We say that X has standard Normal distribution if $\mu = 0$ & $\sigma = 1$

[Standard Normal Distribution denoted as Z]

- The PDF and CDF of Standard Normal

$\phi(z)$ $\Phi(z)$

PDF

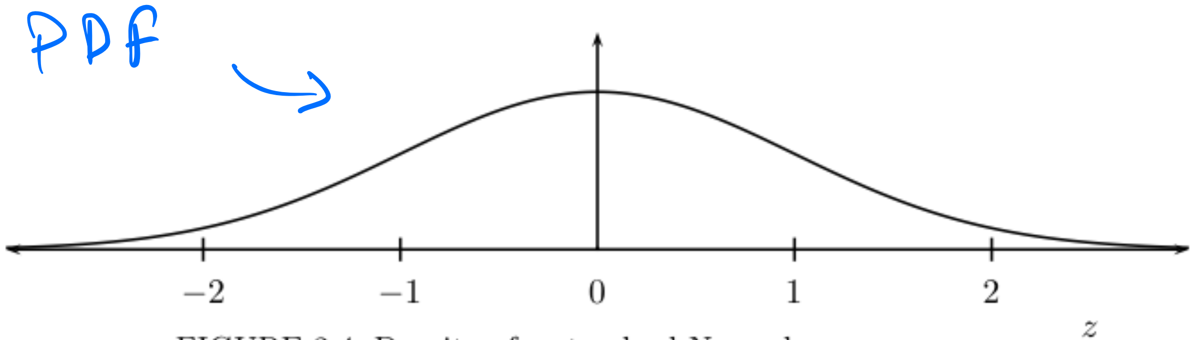


FIGURE 2.4. Density of a standard Normal.

Useful facts :

(i) If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

(ii) If $Z \sim N(0, 1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

(iii) If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$ are independent, then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

It follows from (i) that if $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

* Thus we can compute any probabilities we want as long as we can compute the CDF $\Phi(z)$ of standard Normal

eg Suppose $X \sim N(3,5)$. Find $\mathbb{P}(X>1)$

$X \sim N(\mu,\sigma^2) \Rightarrow \mu=3 \,;\, \sigma^2=5$

- $\mathbb{P}(X>1) = 1-\mathbb{P}(X<1)$

$$= 1-\mathbb{P}\left(Z<\frac{x-\mu}{\sigma}\right)$$

$$= 1-\mathbb{P}\left(Z<\frac{1-3}{\sqrt{5}}\right)$$

$$= 1-\Phi(-0.8944) = 0.81$$

- Now find $q = \Phi^{-1}(0.2)$

$\therefore$ Need to find q such that

$\mathbb{P}(X<q) = 0.2$

$\therefore$ $0.2 = \mathbb{P}(X<q) = \mathbb{P}\left(Z<\frac{q-\mu}{\sigma}\right) = \Phi\left(\frac{q-\mu}{\sigma}\right)$

\# From Normal tables $\Phi(-0.8416) = 0.2$

$\therefore$ $-0.8416 = \frac{q-\mu}{\sigma} = \frac{q-3}{\sqrt{5}} \Rightarrow \underline{\underline{q = 1.1181}}$

• EXPONENTIAL DISTRIBUTION

X has an Exponential distribution with parameter β, written as $X \sim Exp(\beta)$
if

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

where $\beta > 0$.

- The exponential distribution is used to model the lifetimes of electronic components and the waiting times between rare events

• GAMMA DISTRIBUTION

For $a > 0$, the Gamma function is defined by

$$\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy$$

parameters: a, β

written as $X \sim Gamma(a, \beta)$

$$f(x) = \frac{1}{\beta^a \Gamma(a)} x^{a-1} e^{-x/\beta}, \quad x > 0$$

where $a, \beta > 0$

→ The exponential distribution is just a Gamma$(1, \beta)$ distribution

If $X_i \sim$ Gamma(a_i, β) are independent

the $\sum_{i=1}^{n} X_i \sim$ Gamma $\left(\sum_{i=1}^{n} a_i, \beta\right)$

· BETA DISTRIBUTION

X has a Beta distribution, with parameters $a > 0$, $\beta > 0$, written as $X \sim \text{Beta}(a, \beta)$

if

$$f(x) = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} x^{a-1}(1-x)^{\beta-1} \quad ; \; 0 < x < 1$$

· t and CAUCHY DISTRIBUTION

X has a t-distribution with v degrees of freedom; written as $X \sim t_v$ if

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v/2)} \frac{1}{\left(1+\frac{x^2}{v}\right)^{(v+1)/2}}$$

Note: t is similar to Normal but has thicker tails.

↳ Normal corresponds to a t with $v = \infty$

- The Cauchy distribution is a special case of the t distribution [v = 1]

$$f(x) = \frac{1}{\pi(1+x^2)}.$$ → density

To see that this is indeed a density: (proof)

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d\tan^{-1}(x)}{dx} \\
&= \frac{1}{\pi}\left[\tan^{-1}(\infty) - \tan^{-1}(-\infty)\right] = \frac{1}{\pi}\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1.
\end{aligned}$$

- χ^2 DISTRIBUTION

X has a χ^2 distribution with p degrees of freedom - written as $X \sim \chi^2_p$, if

$$f(x) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad x > 0$$

Note: If $Z_1, \dots, Z_p$ are independent standard Normal variables then

$$\sum_{i=1}^{p} Z_i^2 \sim \chi^2_p$$

2.5 - Bivariate Distributions

2.19 Definition. *In the continuous case, we call a function $f(x,y)$ a* PDF *for the random variables (X,Y) if*

(i) $f(x,y) \geq 0$ for all (x,y),

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$ and,

(iii) for any set $A \subset \mathbb{R}\times\mathbb{R}$, $\mathbb{P}((X,Y)\in A) = \int\int_A f(x,y)dxdy$.

In the discrete or continuous case we define the joint CDF as $F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$.

* Note: Given a pair of discrete random variables X & Y, define joint mass function by $\underbrace{f(x,y)}_{f_{X,Y}} = \mathbb{P}(X = x, Y = y)$

eg

Let (X, Y) have density

$$f(x,y) = \begin{cases} x+y, & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Then

$$\int_0^1 \int_0^1 (x+y)\,dx\,dy = \int_0^1 \left[\int_0^1 x\,dx\right] dy + \int_0^1 \left[\int_0^1 y\,dx\right] dy$$

$$= \int_0^1 \frac{1}{2}\,dy + \int_0^1 y\,dy = \frac{1}{2} + \frac{1}{2} = 1$$

$\therefore$ Verifies that this is a PDF

Nice example

2.22 Example. If the distribution is defined over a non-rectangular region, then the calculations are a bit more complicated. Here is an example which I borrowed from DeGroot and Schervish (2002). Let (X, Y) have density

$$f(x,y) = \begin{cases} c\,x^2 y & \text{if } x^2 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note first that $-1 \le x \le 1$. Now let us find the value of c. The trick here is to be careful about the range of integration. We pick one variable, x say, and let it range over its values. Then, for each fixed value of x, we let y vary over its range, which is $x^2 \le y \le 1$. It may help if you look at Figure 2.5. Thus,

$$\begin{aligned} 1 &= \int\!\!\int f(x,y)\,dy\,dx = c\int_{-1}^{1}\int_{x^2}^{1} x^2 y\,dy\,dx \\ &= c\int_{-1}^{1} x^2\left[\int_{x^2}^{1} y\,dy\right]dx = c\int_{-1}^{1} x^2\,\frac{1-x^4}{2}\,dx = \frac{4c}{21}. \end{aligned}$$

$\frac{y^2}{2} = \frac{1}{2} - \frac{x^4}{2} = \frac{1-x^4}{2}$

$1 = \frac{4c}{21}$

$c = \frac{21}{4}$

Hence, $c = 21/4$. Now let us compute $\mathbb{P}(X \ge Y)$. This corresponds to the set $A = \{(x,y);\, 0 \le x \le 1,\, x^2 \le y \le x\}$. (You can see this by drawing a diagram.) So,

$$\begin{aligned} \mathbb{P}(X \ge Y) &= \frac{21}{4}\int_0^1\int_{x^2}^{x} x^2 y\,dy\,dx = \frac{21}{4}\int_0^1 x^2\left[\int_{x^2}^{x} y\,dy\right]dx \\ &= \frac{21}{4}\int_0^1 x^2\left(\frac{x^2-x^4}{2}\right)dx = \frac{3}{20}. \quad \blacksquare \end{aligned}$$

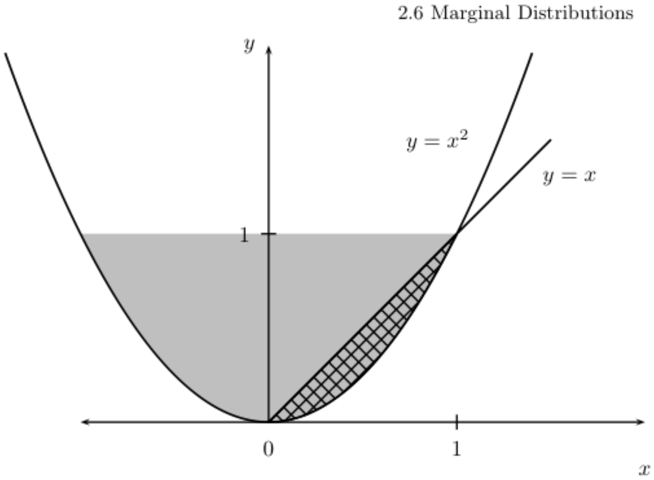


FIGURE 2.5. The light shaded region is $x^2 \le y \le 1$. The density is positive over this region. The hatched region is the event $X \ge Y$ intersected with $x^2 \le y \le 1$.

2.6 – Marginal Distributions

2.23 Definition. *If (X, Y) have joint distribution with mass function $f_{X,Y}$, then the* **marginal mass function for** X *is defined by*

$$f_X(x) = \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f(x, y) \tag{2.4}$$

and the **marginal mass function for** Y *is defined by*

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y) = \sum_x f(x, y). \tag{2.5}$$

eg Suppose that $f_{X,Y}$ is given by

	$Y=0$	$Y=1$	
$X=0$	$1/10$	$2/10$	$3/10$
$X=1$	$3/10$	$4/10$	$7/10$
	$4/10$	$6/10$	

→ The marginal distribution for X corresponds to the row totals

→ The marginal distribution for Y corresponds to the columns total

Definition: For continuous random variable the margin densities are

$$f_X(x) = \int f(x,y)\,dy \quad \&$$

$$f_Y(y) = \int f(x,y)\,dx$$

→ the corresponding marginal distribution functions are denoted by F_X & F_Y

eg Suppose that

$$f_{X,Y}(x,y) = e^{-(x+y)}$$

for $x, y \geq 0$ → then

$$f_X(x) = \int_0^\infty e^{-y}\,dy = e^{-x}$$

eg2

Suppose that

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

then

$$f_Y(y) = \int_0^1 (x+y)\,dx = \int_0^1 x\,dx + \int_0^1 y\,dx$$

$$= \left[\frac{x^2}{2}\right]_0^1 + y = \underline{\frac{1}{2} + y}$$

eg3 Let (X, Y) have density

$$f(x,y) = \begin{cases} \frac{21}{4}x^2 y & \text{if } x^2 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$

$$f_X(x) = \int f(x,y)\,dy = \frac{21}{4}x^2 \int_{x^2}^1 y\,dy$$

$$= \frac{21}{8}x^2(1-x^4) \text{ for } -1 \le x \le 1$$

$f_X(x) = 0$ otherwise

2.7 - Independent Random Variables

2.29 Definition. *Two random variables X and Y are* **independent** *if, for every A and B,*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \tag{2.7}$$

and we write $X \amalg Y$. Otherwise we say that X and Y are **dependent** *and we write $X \not\amalg Y$.*

* To check if 2 random variables are independent we need to check equation [2.7] found above for all subsets A & B.

Note:

Let X & Y have joint PDF $f_{X,Y}$

then $X \amalg Y$ if and only if

$f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all values of x & y

eg

2.31 Example. Let X and Y have the following distribution:

	$Y=0$	$Y=1$	
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
	1/2	1/2	1

Then, $f_X(0) = f_X(1) = 1/2$ and $f_Y(0) = f_Y(1) = 1/2$. X and Y are independent because $f_X(0)f_Y(0) = f(0,0)$, $f_X(0)f_Y(1) = f(0,1)$, $f_X(1)f_Y(0) = f(1,0)$, $f_X(1)f_Y(1) = f(1,1)$. Suppose instead that X and Y have the following distribution:

	$Y=0$	$Y=1$	
X=0	1/2	0	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

These are not independent because $f_X(0)f_Y(1) = (1/2)(1/2) = 1/4$ yet $f(0,1) = 0$. ∎

eg2

suppose that X & Y are independent
& both have the same density

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore f(x,y) = f_X(x) f_Y(y) = \begin{cases} 4xy & 0 \le y \le 1,\ 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore P(X+Y\le 1) = \iint_{x+y\le 1} f(x,y)\,dy\,dx$$

$$= 4\int_0^1 x\left[\int_0^{1-x} y\,dy\right]dx$$

$$= 4\int_0^1 x\frac{(1-x^2)}{2}\,dx = \frac{1}{6}$$

<u>Note</u>: Verifying independence

Suppose that the range of X & Y is a (possibly infinite) rectangle.

If $f(x,y) = g(x)h(y)$ for some functions g & h (not necessarily probability density function) then X & Y are independent

eg

Let X & Y have density

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & \text{if } x>0 \text{ \& } y>0 \\ 0 & \text{otherwise} \end{cases}$$

The range of X & Y is a rectangle

$(0,\infty) \times (0,\infty)$

we can write $f(x,y) = g(x)h(y)$

where $g(x) = 2e^{-x}$ & $h(y) = e^{-2y}$

$\therefore X \perp\!\!\!\perp Y$ $\therefore$ independent

2.8 - Conditional Distributions

- If X & Y are discrete, then we can compute the conditional distribution of X, given that we have observed $Y=y$

$$* \mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$$

2.35 Definition. *The* **conditional probability mass function** *is*

$$f_{X|Y}(x|y) = \mathbb{P}(X=x|Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

if $f_Y(y) > 0$.

Note:

For continuous distributions, we use the same definitions, but interpretation differs.

→ Discrete: $f_{X|Y}(x|y) = \mathbb{P}(X=x \mid Y=y)$

→ Continuous: need to integrate to get a probability.

2.36 Definition. *For continuous random variables, the* **conditional probability density function** *is*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

assuming that $f_Y(y) > 0$. *Then,*

$$\mathbb{P}(X \in A | Y = y) = \int_A f_{X|Y}(x|y)dx.$$

eg

Let X & Y have joint uniform distribution on the unit square.

Thus $f_{X|Y}(x|y) = 1$ for $0 \leq x \leq 1$.

0 otherwise

Given that $Y = y$, X is uniform $(0,1)$

we can write that $X|Y \sim \text{Uniform}(0,1)$

→ from the definition of conditional density

$$f_{X,Y}(x,y) = f_{X|Y}(y|x) f_Y(y)$$

$$= f_{Y|X}(y|x) f_X(x)$$

eg 2 let

$$f(x,y)=\begin{cases} x+y & 0\leq x\leq 1,\ 0\leq y\leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{P}(X<1/4 \mid Y=1/3)$; $f_Y(y)=y+\frac{1}{2}$

$$\therefore\ f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}=\frac{x+y}{y+1/2}$$

so

$$\mathbb{P}(X<1/4 \mid Y=1/3)=\int_0^{1/4} f_{X|Y}(x \mid 1/3)\,dx$$

$$=\int_0^{1/4}\frac{x+1/3}{1/3+\frac{1}{2}}\,dx=\frac{1/32+1/12}{1/3+1/2}=\frac{11}{80}\ //$$

2.39 Example. Suppose that $X \sim \text{Uniform}(0,1)$. After obtaining a value of X we generate $Y|X = x \sim \text{Uniform}(x, 1)$. What is the marginal distribution

---

avoid this problem by defining things in terms of the PDF. The fact that this leads to a well-defined theory is proved in more advanced courses. Here, we simply take it as a definition.

of Y? First note that,

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$\frac{1}{b-a}$

and

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

So,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal for Y is

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)dx = \int_0^y \frac{dx}{1-x} = -\int_1^{1-y} \frac{du}{u} = -\log(1-y)$$

for $0 < y < 1$. ■

2.40 Example. Consider the density in Example 2.28. Let's find $f_{Y|X}(y|x)$. When $X = x$, y must satisfy $x^2 \le y \le 1$. Earlier, we saw that $f_X(x) = (21/8)x^2(1-x^4)$. Hence, for $x^2 \le y \le 1$,

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} = \frac{2y}{1-x^4}.$$

Now let us compute $\mathbb{P}(Y \ge 3/4 | X = 1/2)$. This can be done by first noting that $f_{Y|X}(y|1/2) = 32y/15$. Thus,

$$\mathbb{P}(Y \ge 3/4 | X = 1/2) = \int_{3/4}^1 f(y|1/2)dy = \int_{3/4}^1 \frac{32y}{15}dy = \frac{7}{15}. \quad ■$$

2.9 - Multivariate Distributions & IID Samples

- Let $X = (X_1, \dots, X_n)$ where $X_1 \dots X_n$ are random variables

→ We call X random vector

- Let $f(x_1, \dots, x_n)$ denote the PDF

↳ define their marginals, conditionals...

We say that $X_1, \dots, X_n$ are independent, if for every $A_1, \dots, A_n$

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

⇝ it suffices to check that

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

2.41 Definition. *If* $X_1, \ldots, X_n$ *are independent and each has the same marginal distribution with* CDF F, *we say that* $X_1, \ldots, X_n$ *are* IID *(independent and identically distributed) and we write*

$$X_1, \ldots X_n \sim F.$$

If F *has density* f *we also write* $X_1, \ldots X_n \sim f$. *We also call* $X_1, \ldots, X_n$ *a* **random sample of size** n **from** F**.**

2.10 - 2 Important Multivariate Distributions

• MULTINOMIAL

→ Multivariate version of Binomial.

• Consider drawing a ball from an urn which has balls with $\underline{k}$ different colours → labelled "color 1, color 2 ..., color k"

Let $p = (p_1, \dots, p_k)$ where $p_j \geq 0$

$\sum_{j=1}^{k} p_j = 1$; p_j = probability of drawing a ball of color j.

→ Draw $\underset{\sim}{n}$ times (independent draws with replacement)

and let $X = (X_1, \dots, X_k)$ where

* X_j = # of times that color j appears

Hence, $n = \sum_{j=1}^{k} x_j$

We say that X has a Multinomial (n, p) distribution.

$\leadsto$ written as $X \sim \text{Multinomial}(n, p)$

The probability function is:

$$f(x) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}$$

where

$$\binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1! \cdots x_k!}$$

⊛ Lemma: Suppose $X \sim \text{Multinomial}(n, p)$ where $X = (X_1 \ldots X_k)$ & $p = (p_1, \ldots, p_k)$

The marginal distribution $X_j \sim \text{Bin}(n, p_j)$

• MULTIVARIATE NORMAL

→ The univariate Normal has 2 parameters
(μ, σ)

* In multivariate version, μ is a vector
& σ is replaced by matrix Σ

• To begin, let

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$$

where $Z_1, \ldots, Z_k \sim N(0,1)$ independent.

↳ The density of Z is:

$$\begin{aligned} f(z) &= \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{k} z_j^2\right\} \\ &= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} z^T z\right\}. \end{aligned}$$

* Note: if a & b are vectors then
$a^T b = \sum_{i=1}^{k} a_i b_i$.

• We say that Z has a standard multivariate Normal distribution

written as $Z \sim N(0, I)$

* were 0 represents a vector of k zeroes

& I is $k \times k$ identity matrix.

→ More generally, a vector X has a multivariate Normal distribution $X \sim N(\mu, \Sigma)$, if it has density.

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2}|(\Sigma)|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \tag{2.10}$$

where $|\Sigma|$ denotes the determinant of Σ, μ is a vector of length k and Σ is a $k \times k$ symmetric, positive definite matrix. [9] Setting $\mu = 0$ and $\Sigma = I$ gives back the standard Normal.

Since Σ is symmetric and positive definite, it can be shown that there exists a matrix $\Sigma^{1/2}$ — called the square root of Σ — with the following properties: (i) $\Sigma^{1/2}$ is symmetric, (ii) $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ and (iii) $\Sigma^{1/2}\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2} = I$ where $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.

[8] Σ^{-1} is the inverse of the matrix Σ.

[9] A matrix Σ is positive definite if, for all nonzero vectors x, $x^T\Sigma x > 0$.

Theorem (1)

If $Z \sim N(0, I)$ & $X = \mu + \Sigma^{1/2} Z$

then $X \sim N(\mu, \Sigma)$

Conversely, if $X \sim N(\mu, \Sigma)$ then

$\Sigma^{-1/2}(X - \mu) \sim N(0, I)$

- Suppose we partition a random Normal vector X as $X = (X_a, X_b)$.

we can similarly partition $\mu = (\mu_a, \mu_b)$

&

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Theorem (2)

2.44 Theorem. *Let* $X \sim N(\mu, \Sigma)$. *Then:*

(1) *The marginal distribution of* X_a *is* $X_a \sim N(\mu_a, \Sigma_{aa})$.

(2) *The conditional distribution of* X_b *given* $X_a = x_a$ *is*

$$X_b | X_a = x_a \sim N\left(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a),\ \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \right).$$

(3) *If* a *is a vector then* $a^T X \sim N(a^T \mu, a^T \Sigma a)$.

(4) $V = (X - \mu)^T \Sigma^{-1} (X - \mu) \sim \chi^2_k$.

2.11 - Transformations of Random Variables

- Suppose that X is a random variable with PDF f_X & CDF F_X

Let $Y = r(X)$ be a function of X

eg $Y = X^2$ or $Y = e^X$

→ We call $Y = r(X)$ a transformation of X.

\# How do we compute PDF & CDF of Y?

- DISCRETE CASE

The mass function of Y is given by

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(r(X) = y)$$

$$= \mathbb{P}(\{x ; r(x) = y\}) = \mathbb{P}(X \in r^{-1}(y))$$

eg

2.45 Example. Suppose that $\mathbb{P}(X=-1)=\mathbb{P}(X=1)=1/4$ and $\mathbb{P}(X=0)=1/2$. Let $Y=X^2$. Then, $\mathbb{P}(Y=0)=\mathbb{P}(X=0)=1/2$ and $\mathbb{P}(Y=1)=\mathbb{P}(X=1)+\mathbb{P}(X=-1)=1/2$. Summarizing:

x	$f_X(x)$
-1	1/4
0	1/2
1	1/4

y	$f_Y(y)$
0	1/2
1	1/2

Y takes fewer values than X because the transformation is not one-to-one. ■

• CONTINUOUS CASE

3 steps for finding f_Y:

Three Steps for Transformations

1. For each y, find the set $A_y = \{x: \ r(x) \le y\}$.

2. Find the CDF

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(r(X) \le y) \\ &= \mathbb{P}(\{x;\ r(x) \le y\}) \\ &= \int_{A_y} f_X(x)dx. \end{aligned} \tag{2.11}$$

3. The PDF is $f_Y(y) = F'_Y(y)$.

eg 1

Let $f_X(x) = e^{-x}$ for $x > 0$

$F_X(x) = \int_0^x f_X(s)\,ds = 1 - e^{-x}$

Let $Y = r(X) = \log X$

$\cdot A_y = \{ x : x \le \underbrace{e^y}_{y} \}$

$\therefore F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(r(X) \le y)$

$= \mathbb{P}(\log X \le y)$

$= \mathbb{P}(X \le e^y) = F_X(e^y) = 1 - e^{-e^y}$

$\therefore f_Y(y) = e^y e^{-e^y}$ for $y \in \mathbb{R}$

eg 2

2.47 Example. Let $X \sim \text{Uniform}(-1, 3)$. Find the PDF of $Y = X^2$. The density of X is

$$f_X(x) = \begin{cases} 1/4 & \text{if } -1 < x < 3 \\ 0 & \text{otherwise.} \end{cases}$$

$(-1)^2 = 1$
$(3)^2 = 9$

Y can only take values in $(0, 9)$. Consider two cases: (i) $0 < y < 1$ and (ii) $1 \leq y < 9$. For case (i), $A_y = [-\sqrt{y}, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/2)\sqrt{y}$. For case (ii), $A_y = [-1, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/4)(\sqrt{y} + 1)$. Differentiating F we get

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & \text{if } 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & \text{if } 1 < y < 9 \\ 0 & \text{otherwise.} \end{cases} \blacksquare$$

When r is strictly monotone increasing or strictly monotone decreasing then r has an inverse $s = r^{-1}$ and in this case one can show that

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|. \tag{2.12}$$

2.12 - Transformations of Several Random variables

- For example if X & Y are given random variables.

⇝ we might want to know the distribution of X/Y, $X+Y$, $\max\{X,Y\}$ or $\min\{X,Y\}$

⇝ let $Z = r(X,Y)$

Steps of finding f_Z

Three Steps for Transformations

1. For each z, find the set $A_z = \{(x,y) : r(x,y) \le z\}$.
2. Find the CDF

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) = \mathbb{P}(r(X,Y) \le z) \\ &= \mathbb{P}(\{(x,y);\ r(x,y) \le z\}) = \int\int_{A_z} f_{X,Y}(x,y)\, dx\, dy. \end{aligned}$$

3. Then $f_Z(z) = F'_Z(z)$.

eg

Let $X_1, X_2 \sim \text{Uniform}(0,1)$ be independent.

Find the density of $Y = X_1 + X_2$

$\to$ The joint density of (X_1, X_2) is

$$f(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $r(x_1, x_2) = x_1 + x_2$. Now

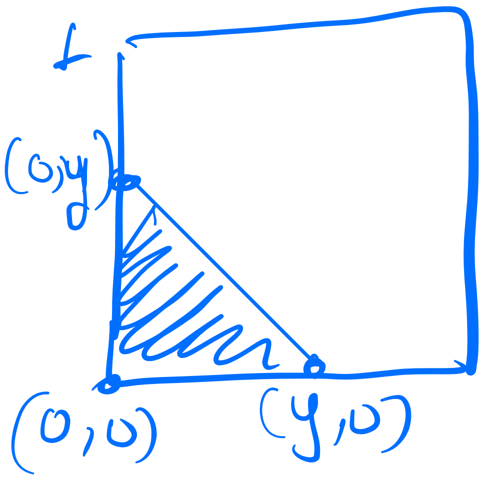
$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(r(X_1, X_2) \leq y)$$

$$= \mathbb{P}(\{(x_1, x_2) : r(x_1, x_2) \leq y\})$$

$$= \iint_{A_y} f(x_1, x_2)\, dx_1\, dx_2$$

* Finding A_y *

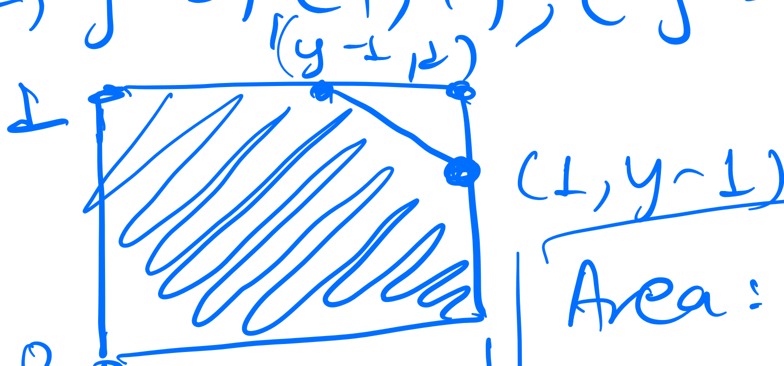
- Suppose $0 < y \leq 1$. Then A_y is the triangle with vertices $(0,0)$ $(y,0)$ & $(0,y)$



$$\iint_{A_y} f(x_1,x_2)\,dx_1\,dx_2$$

$\hookrightarrow$ $\boxed{\text{area of } \Delta = y^2/2}$

$\rightarrow$ if $1 < y < 2$, then A_y is everything in the unit square except the triangle $\rightarrow$ with vertices $(1, y-1)$, $(1,1)$, $(y-1, 1)$



$$\boxed{\text{Area} : 1 - \frac{(2-y)^2}{2}}$$

$\therefore$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \dfrac{y^2}{2} & 0 \le y \le 1 \\ 1 - \dfrac{(2-y)^2}{2} & 1 \le y < 2 \\ 1 & y \geq 2 \end{cases}$$

By differentiation, the PDF

$$f_Y(y) = \begin{cases} y & 0 \le y < 1 \\ 2-y & 1 \le y \le 2 \\ 0 & \text{otherwise.} \end{cases}$$